

# Query Form

SREP		
Journal:	41598	[Art. ID: 84969]

## Journal: SREP

**Author:** The following queries have arisen during the editing of your manuscript. Please answer queries by making the requisite corrections at the appropriate positions in the text.

Query	Details Required	Author's Response
<a href="#">AQ1</a>	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.	
<a href="#">AQ2</a>	Kindly check and confirm the city name in affiliation 6 and 7.	
<a href="#">AQ3</a>	Kindly check and confirm thousand separator in values 4,988,755.	



OPEN

# Investigating the microbial ecology of coastal hotspots of marine nitrogen fixation in the western North Atlantic

Seaver Wang<sup>1</sup>, Weiyi Tang<sup>2</sup>, Erwan Delage<sup>3</sup>, Scott Gifford<sup>4</sup>, Hannah Whitby<sup>5</sup>, Aridane G. González<sup>6,7</sup>, Damien Eveillard<sup>3</sup>, Hélène Planquette<sup>7</sup> & Nicolas Cassar<sup>1,7</sup>✉

Variation in the microbial cycling of nutrients and carbon in the ocean is an emergent property of complex planktonic communities. While recent findings have considerably expanded our understanding of the diversity and distribution of nitrogen (N<sub>2</sub>) fixing marine diazotrophs, knowledge gaps remain regarding ecological interactions between diazotrophs and other community members. Using quantitative 16S and 18S V4 rDNA amplicon sequencing, we surveyed eukaryotic and prokaryotic microbial communities from samples collected in August 2016 and 2017 across the Western North Atlantic. Leveraging and significantly expanding an earlier published 2015 molecular dataset, we examined microbial community structure and ecological co-occurrence relationships associated with intense hotspots of N<sub>2</sub> fixation previously reported at sites off the Southern New England Shelf and Mid-Atlantic Bight. Overall, we observed a negative relationship between eukaryotic diversity and both N<sub>2</sub> fixation and net community production (NCP). Maximum N<sub>2</sub> fixation rates occurred at sites with high abundances of mixotrophic stramenopiles, notably *Chrysophyceae*. Network analysis revealed such stramenopiles to be keystone taxa alongside the haptophyte diazotroph host *Braarudosphaera bigelowii* and chlorophytes. Our findings highlight an intriguing relationship between marine stramenopiles and high N<sub>2</sub> fixation coastal sites.

Marine phytoplankton mediate significant fluxes of carbon, oxygen, nitrogen, and other elements between organic, atmospheric, and oceanic pools. Nitrogen (N<sub>2</sub>) fixation is a critical biogeochemical process in which specialized marine prokaryotes provide the surface planktonic community with a supply of new nitrogen obtained from the atmosphere. N<sub>2</sub> fixation supports marine primary production and can increase the sequestration of biological carbon in the deep ocean via sinking plankton biomass<sup>1</sup>. As a result, marine N<sub>2</sub> fixation is an important process influencing the ocean carbon cycle, with implications for marine carbon sink strength and global climate<sup>1–3</sup>.

Marine N<sub>2</sub> fixation and the diazotrophic microbes that perform this function are strongly influenced by nutrient availability, ecosystem conditions, and the ecology and structure of the surface microbial community<sup>4–6</sup>. Consequently, microbial oceanographers are devoting considerable effort to studying the links between N<sub>2</sub> fixation, the physical and chemical environment, and community structure. While studies of bottom-up controls on N<sub>2</sub> fixation are well underway<sup>5,7–9</sup>, the complex ecological relationships governing diazotrophs remain only partially explored. For example, there is increasing field evidence regarding the significance of top-down factors in controlling diazotroph populations<sup>10–12</sup>, potentially influencing the global distribution of diazotrophs<sup>13</sup>. Our evolving understanding of such ecological dynamics stems from discoveries over the last decade that have significantly altered the scientific knowledge behind marine N<sub>2</sub> fixation. Researchers have discovered a number of additional marine diazotrophs, including UCYN-A—a cyanobacterial endosymbiont of the haptophyte phytoplankton

<sup>1</sup>Division of Earth and Ocean Sciences, Duke University, Grainger Environment Hall, 9 Circuit Drive, Box 90328, Durham, NC 27708, USA. <sup>2</sup>Department of Geosciences, Princeton University, Princeton, NJ, USA. <sup>3</sup>LS2N, UMR 6004, CNRS, Université de Nantes, 44000 Nantes, France. <sup>4</sup>Department of Marine Sciences, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. <sup>5</sup>Department of Earth, Ocean, and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, UK. <sup>6</sup>Instituto de Oceanografía y Cambio Global (IOCG), Universidad de Las Palmas de Gran Canaria, ULPGC, Las Palmas, Spain. <sup>7</sup>Laboratoire des Sciences de l'Environnement Marin (LEMAR), Institut Universitaire Européen de la Mer (IUEM), Technopôle Brest-Iroise, 13 Plouzané, 29280 Locmaria-Plouzané, France. ✉email: nicolas.cassar@duke.edu

AQ1 AQ2

*Braarudosphaera bigelowii*<sup>14,15</sup>—and numerous groups of non-cyanobacterial diazotrophs (NCDs), including members of Gammaproteobacteria and Planctomycetes<sup>10,16–18</sup>. Other key discoveries include the finding that active marine N<sub>2</sub> fixation can occur in the presence of nitrate or ammonium<sup>19–21</sup>, that coastal rates of N<sub>2</sub> fixation are much higher than previously thought<sup>22,23</sup>, and that N<sub>2</sub> fixation can be significant even in cold high-latitude regions<sup>20,24,25</sup>, as reviewed in Zehr and Capone, 2020<sup>5</sup>.

Such revelations also make it clear that the surrounding microbial community likely exerts ecological influences over diazotrophs in addition to the physical environment and nutrient regime. Colony-forming or host-associated N<sub>2</sub> fixing bacteria may interact with other marine microbes inhabiting colony or host cell surfaces<sup>26–28</sup>. Community interactions might also provide diazotrophs with sources of organic-rich particles and aggregates<sup>29</sup>, and recycled essential nutrients such as phosphorus<sup>30</sup>. Grazers<sup>11,31</sup>, viruses<sup>32</sup>, and parasites represent other ways in which ecological dynamics might regulate N<sub>2</sub> fixing organisms and their hosts via predation upon the host algal cells. Finally, the marine environment itself can be altered by local microbial ecology, such as the transmissivity of the water column to light<sup>33</sup> or the presence of inhibitory compounds produced by competitors, particularly in more restricted coastal waters or at very high cell densities<sup>34</sup>. Marine microbial community structure thus potentially moderates N<sub>2</sub> fixation activity<sup>35</sup>. Hotspots of N<sub>2</sub> fixation might be considered an emergent property of the whole marine microbial community. Such possibilities add considerable value to uncovering ecological conditions or relationships between marine diazotrophs and other taxa within the surface ocean community. While conclusively linking N<sub>2</sub> fixation to microbial abundances and diversity requires understanding underlying ecological mechanisms, we remain at a stage where inferences from the field are necessary to highlight patterns of interest for focused study.

From August 2015 to August 2017, a successive series of research expeditions in the western North Atlantic aboard the *R/V Atlantic Explorer* sought to pair underway measurements of N<sub>2</sub> fixation with sampling of diazotroph and marine microbial community structure, biogeochemical properties, and a survey of net community production rates. Wang et al., 2018 described a marine microbial community dominated by *Aureococcus anophagefferens* associated with a coastal site in the Mid-Atlantic Bight exhibiting elevated net community production rates<sup>36</sup>. Using a new method for high-resolution underway observations of N<sub>2</sub> fixation, Tang et al., 2019 identified these productive coastal waters of the Mid-Atlantic Bight and Southern New England Shelf also to be hotspots of elevated N<sub>2</sub> fixation dominated by the cyanobacterial diazotroph UCYN-A, with extremely high rates up to 167  $\mu\text{mol N m}^{-3} \text{ day}^{-1}$  measured off the New Jersey coast<sup>22</sup>. Tang et al. 2020 analyzed macronutrient and micronutrient concentrations within coastal ecosystems with high rates of N<sub>2</sub> fixation. They discovered these environments to be relatively replete in iron and nitrate, identifying phosphorus and temperature as key factors governing N<sub>2</sub> fixation activity and diazotrophic community structure, respectively<sup>37</sup>. In the same study, a *nifH* molecular survey identified UCYN-A as dominant in cold, subpolar coastal waters with diazotroph populations shifting towards *Trichodesmium* in warmer, oligotrophic waters offshore.

In this study, we investigated the microbial diversity and ecological relationships underlying the published extensive regional survey of marine N<sub>2</sub> fixation rates in summer reported in Tang et al., 2019 and Tang et al., 2020. Specifically, through statistical inferences, we attempted to assess potential competitive interactions associated with previously-identified coastal hotspots of N<sub>2</sub> fixation, refining our understanding of the ecological niche occupied by diazotrophic plankton in this ecosystem. We hypothesize that given previous findings of record UCYN-A abundances, network-based statistical analysis of co-occurrence patterns will reveal that UCYN-A's host *B. bigelowii* plays a disproportionately crucial ecological role within this productive coastal community, potentially reflecting its role in transferring bioavailable N into the microbial environment.

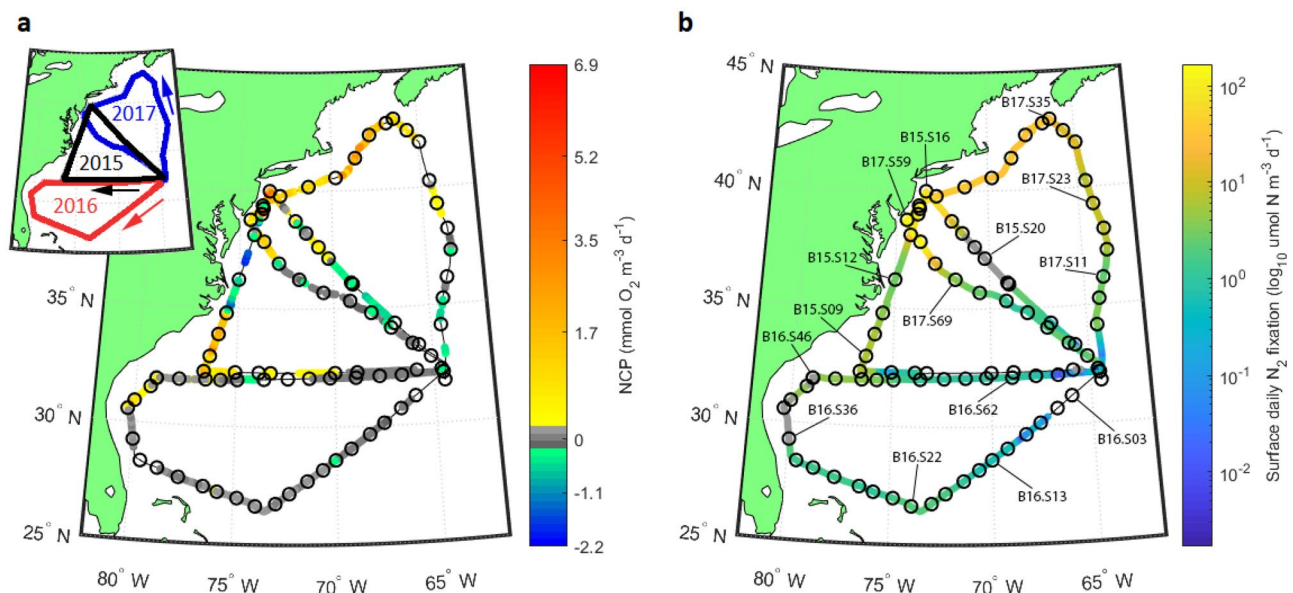
## Results and discussion

**Microbial ecology: marine microbial community composition and spatial and abundance patterns.** The earlier discovery that UCYN-A is most likely responsible for observed coastal peaks in N<sub>2</sub> fixation in the Mid-Atlantic Bight and Southern New England Shelf<sup>22</sup> is intriguing given that UCYN-A is presumed to be an obligate endosymbiont of its host *Braarudosphaera bigelowii*. UCYN-A's record abundances (up to  $4 \times 10^7$  *nifH* copies L<sup>-1</sup>)<sup>37</sup> raise the question of what ecological conditions permit its host organism to be competitively successful and whether locally elevated N<sub>2</sub> fixation rates could be partially enabled by the success of the host phytoplankton in addition to the favorability of a niche for diazotrophic N<sub>2</sub> fixation.

In our molecular survey across the western North Atlantic (Fig. 1), marine prokaryotic 16S rDNA samples typically displayed high abundances of SAR11 and *Prochlorococcus* as well as appreciable fractions of Bacterioplankton belonging to the SAR86 clade (*Oceanospirillales*), *Flavobacteriales*, and *Rhodospirillales*. A noticeable shift in community composition towards greater fractions of *Planctomycetes*, *Phycisphaerales*, and *Sphingobacteriales* was observed for stations with high N<sub>2</sub> fixation rates and net community production (NCP) rates (Fig. 2).

The highest N<sub>2</sub> fixation and NCP rates corresponded to samples collected in 2015 and 2017 near the New Jersey coast in the Mid-Atlantic Bight (MAB) and off the Southern New England Shelf (Fig. 1). Over the region studied, previous work<sup>22,36,37</sup> reported surface daily N<sub>2</sub> fixation rates ranging between undetectable values ( $< 0.19 \mu\text{mol N m}^{-3} \text{ day}^{-1}$ , see Cassar et al., 2018<sup>38</sup>) and  $167 \mu\text{mol N m}^{-3} \text{ day}^{-1}$  (mean  $\pm$  sd:  $3 \pm 35.1 \mu\text{mol N m}^{-3} \text{ day}^{-1}$ ) and volumetric NCP rates ranging between  $-2.2$  and  $6.9 \text{ mmol O}_2 \text{ m}^{-3} \text{ day}^{-1}$  ( $0.4 \pm 0.9 \text{ mmol O}_2 \text{ m}^{-3} \text{ day}^{-1}$ ).

Generally, eukaryotic 18S rDNA samples were dominated by dinoflagellate OTUs (Fig. 3), particularly *Syndiniales*—endoparasitic alveolates infecting a wide range of organisms from ciliates to cercozoa, zooplankton, and fish eggs<sup>39–41</sup>. The life cycle of *Syndiniales* also involves a free-living stage in which large numbers of *Syndiniales* are dispersed as dinospores, perhaps partially explaining their high abundances in many amplicon studies<sup>39</sup>. Eukaryotic dinoflagellates belonging to *Gonyaulacales*, *Gymnodiniophycidae*, and *Dinophyceae* appeared at around 5–10% relative abundances across most samples.



**Figure 1.** NCP and  $N_2$  fixation rates. (a) Map of  $O_2/Ar$ -based volumetric NCP rates measured via EIMS across the Western North Atlantic in July–August 2015<sup>36</sup>, 2016, and 2017<sup>37</sup>. Positive rates are represented by the warm portion of the color scale, while negative rates are indicated using a cool color scale. Balanced NCP rates are shown in greyscale tones. Sites of molecular sampling are displayed by open circles. (b) Map of surface daily  $N_2$  fixation rates measured using the FARACAS method<sup>22,37</sup>. Sites of molecular sampling are displayed by open circles, with selected sites labeled. Greyscale points indicate sites at which  $N_2$  fixation rates were below the limit of detection for the FARACAS method ( $0.19 \mu\text{mol N m}^{-3} \text{ day}^{-1}$ ). Inset plot illustrates cruise tracks and directions of travel, color-coded for each year of sampling. Plots created using Matlab 2018b (<https://www.mathworks.com/products/matlab.html>).

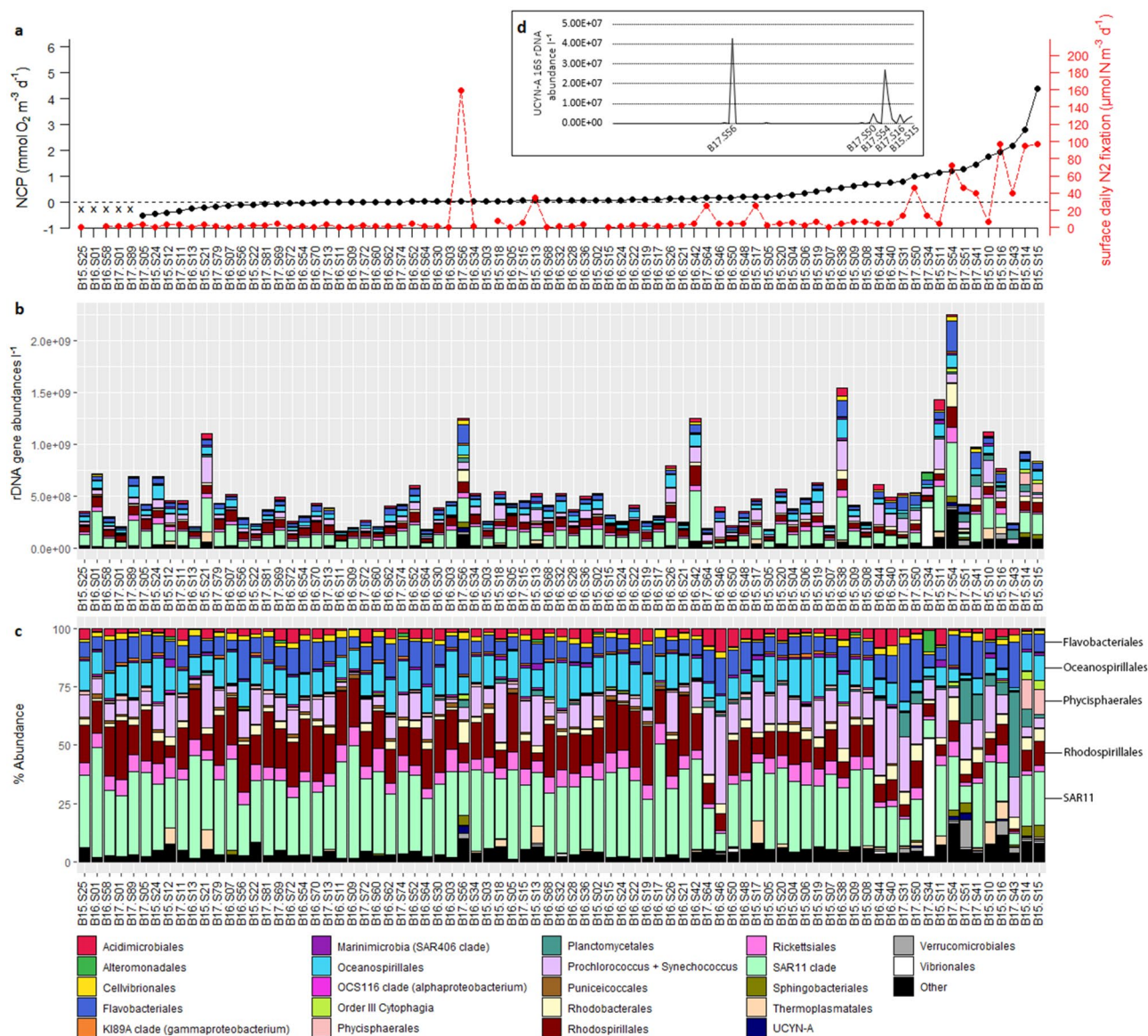
Samples from the MAB and New England shelf exhibiting high  $N_2$  fixation and NCP rates in 2015 and 2017 were dominated by *Chrysophyceae* and the pelagophyte alga *Aureococcus anophagefferens*. Strikingly, of the seven stations with the highest  $N_2$  fixation rates, five showed extremely elevated *Chrysophyceae* abundances (25–55% of 18S rDNA abundances). Several of these samples also displayed elevated abundances of the chlorophyte *Mamiellales*. These samples from the MAB in August 2015 and 2017 clustered closely together during PCoA analysis, suggesting consistency in broad community structure between years (Fig. 4). Analysis of variance testing via the adonis function (999 permutations) in the “vegan” package revealed that the differences in community composition between 18 and 16S samples collected north of the Gulf Stream versus south of the Gulf Stream were statistically significant ( $p < 0.001$ ).

Although we observed high  $N_2$  fixation, high NCP, and high abundances of *Chrysophyceae* in the MAB in both August 2015 and 2017, we note that *Aureococcus anophagefferens* did not bloom here to the same extent in 2017 as in 2015<sup>36</sup>. In 2017, *Aureococcus* occurred at 0.7–10.1% of the eukaryotic community within the MAB, with an estimated  $7.2 \times 10^7$  to  $2.2 \times 10^9$  18S rDNA gene sequences  $l^{-1}$ , contrasted with 2015 sampling ( $1.8 \times 10^8$  to  $2.0 \times 10^{10}$  rDNA genes  $l^{-1}$ ). The absence of a similarly dramatic 2017 *Aureococcus* bloom in the MAB suggests that such events may not recur annually, or that blooms may be subject to variable timing and magnitude. Existing literature reporting upon coastal monitoring of *Aureococcus* indeed demonstrate that *Aureococcus* abundances may fluctuate by an order of magnitude between months during the growing season and that the seasonal biomass peak may precede the start of August by several weeks<sup>43</sup>.

**Broad patterns: community diversity vs.  $N_2$  fixation and NCP.** We observed a negative relationship between  $N_2$  fixation and eukaryotic community Shannon’s H diversity (Pearson =  $-0.73$ ,  $p < 0.001$ ; Spearman’s rho = not significant), in which stations with lower eukaryotic diversity tended to exhibit higher  $N_2$  fixation rates (Fig. 5). A similar negative relationship was observed between NCP rates and eukaryotic diversity (Pearson =  $-0.70$ ,  $p < 0.001$ ; Spearman’s rho =  $-0.31$ ,  $p < 0.05$ ). These correlations appear to be driven by both by decreasing species richness and evenness, as samples drawn from locations with high  $N_2$  fixation and NCP rates tended to contain substantially fewer OTUs, often two standard deviations lower than the rest of the dataset (mean: 397, s.d.: 131), yet were also dominated by particular taxa such as *Chrysophyceae* and *A. anophagefferens*. We note that the linear regressions shown in Fig. 5 are included as visual aids, and not meant to assert that these relationships are necessarily linear.

Prokaryotic Shannon’s H diversity was weakly correlated to  $N_2$  fixation rates (Pearson =  $0.3$ ,  $p < 0.01$ ; Spearman’s rho =  $0.48$ ,  $p < 0.01$ ) but not with NCP ( $p > 0.05$ ) (Fig. 5, Table S2). We also note that samples collected in 2016 typically exhibited a narrow range of eukaryotic diversity compared to the range observed for other years (Fig. 5) as well as generally similar eukaryotic community structure (Figs. 2, 3), whereas the range of prokaryotic diversity was relatively similar across all three years. Diversity patterns were also compared to other parameters





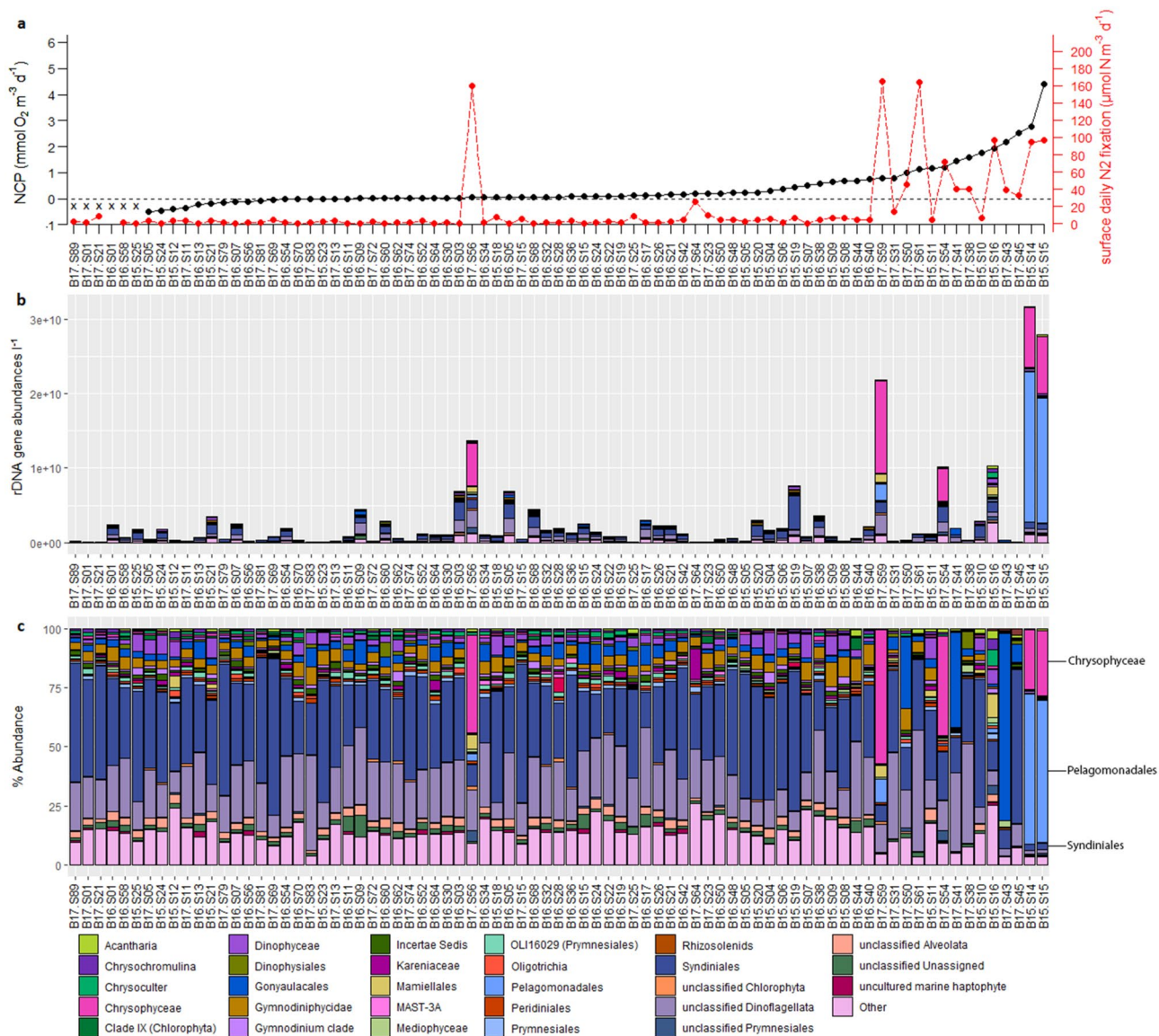
**Figure 2.** Prokaryotic community structure. (a) N<sub>2</sub> fixation rates (red) and NCP rates (black) measured at each sampling site<sup>22,36,37</sup>, with sites ranked from left-to-right in order of increasing NCP. Sites at which NCP measurements were not collected are grouped on the far left and denoted with x-marks. Taxa are organized in alphabetical order from top to bottom, following the order outlined in the legend. (b) Bar plot of 16S prokaryotic taxonomy represented at the order level using quantitative rDNA gene abundances l<sup>-1</sup>. (c) Bar plot of 16S prokaryotic taxonomy at the order level using relative abundances. (d) UCYN-A1 16S rDNA abundance across sampling sites, with sites ranked from left-to-right in order of increasing NCP and selected sites labeled. Figure created in R 3.5.2 (<https://www.r-project.org/>).

including macronutrient concentrations, dissolved and particulate trace metals, and silica, with trends discussed in detail in the Supplementary Material.

These results are of interest for considering the ecological characteristics of the coastal hotspots of N<sub>2</sub> fixation and NCP sampled during this survey. Productivity-diversity relationships in oceanic ecosystems can provide insight into ecological processes involved, such as competitive exclusion<sup>44</sup>, “kill-the-winner”<sup>45</sup>, and co-existence.

Increasing competitive exclusion driven by opportunistic bloom taxa has been proposed as an explanation for the negative slope portion of unimodal productivity-diversity relationships<sup>46</sup>. Earlier studies suggested a unimodal “humped” shape to such relationships, in which diversity increases with productivity, peaking at intermediate levels of production, then falling<sup>47,48</sup>. More recent work has not only documented a range of productivity-diversity relationships ranging from positive to negative to flat<sup>49,50</sup>, but has also raised methodological concerns, postulating instead that diversity may not be correlated with productivity, with rare taxa impacting species richness but not greatly influencing production<sup>51</sup>.

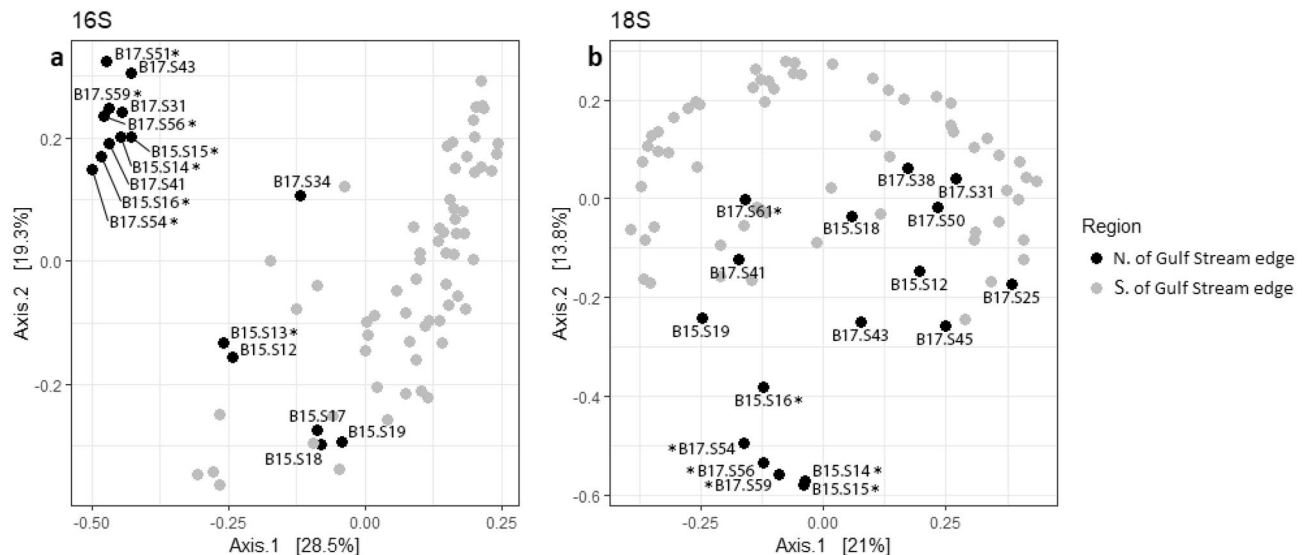
Our results, showing negative relationships with decreasing eukaryotic diversity for increasing N<sub>2</sub> fixation and NCP, raise the possibility that competitive mechanisms could play a role in limiting eukaryotic diversity within



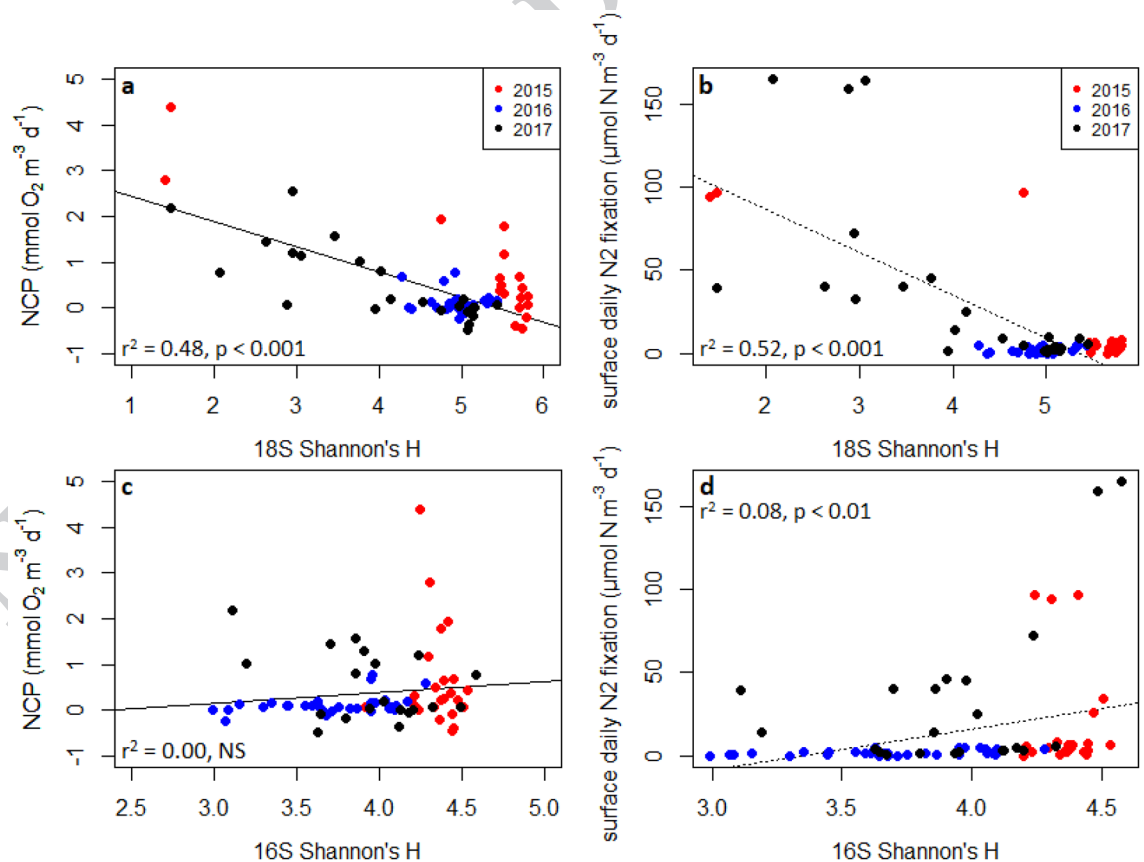
**Figure 3.** Eukaryotic community structure. (a) N<sub>2</sub> fixation rates (red) and NCP rates (black) measured at each sampling site<sup>22,36,37</sup>, with sites ranked from left-to-right in order of increasing NCP. Sites at which NCP measurements were not collected are grouped on the far left and denoted with x-marks. Taxa are organized in alphabetical order from top to bottom, following the order outlined in the legend. (b) Bar plot of 18S eukaryotic taxonomy represented at the 5th taxonomic rank using quantitative rDNA gene abundances  $\text{l}^{-1}$ . (c) Bar plot of 18S eukaryotic taxonomy at the 5th taxonomic rank using relative abundances. Note that *Aureococcus anophagefferens* fall within the order Pelagomonadales, and make up the overwhelming majority of pelagophytes in the sample set. Figure created in R 3.5.2 (<https://www.r-project.org/>).

the sampled coastal blooms. One potential mechanism is that the dominant eukaryotic plankton identified in our sampling, *Chrysophyceae*, and *Aureococcus*, can produce allelopathic compounds to inhibit the growth of algal competitors<sup>34,52,53</sup>. Given the record abundances of UCYN-A observed at these sites<sup>37</sup>, such patterns raise the possibility that UCYN-A and its host *Braarudosphaera bigelowii* can tolerate these chemicals. *Aureococcus* is adapted for low-light conditions and capable of leveraging diverse uptake pathways for dissolved organic carbon and nitrogen<sup>54,55</sup> while *Chrysophyceae* can engulf particulate matter via phagotrophy<sup>56,57</sup>, potentially giving both groups of mixotrophic algae versatile capabilities for out-competing other taxa.

A positive relationship between N<sub>2</sub> fixation and prokaryotic diversity could imply a greater availability of varied bacterial niches within N<sub>2</sub> fixation hotspots, but the observed relationships in this study are too weak to confidently support such a hypothesis. An OTU-based definition for prokaryotic diversity is imperfect, given the ambiguity of the “species” concept for prokaryotes<sup>58,59</sup>. Observed relationships may depend on how diversity is defined and which functional groups (algae, free-living and particle-associated bacteria, viruses) are captured by field methods and considered part of the community. The observed relationships with eukaryotic diversity are also strongly driven by the samples taken in 2017 and 2015 off the New England Shelf, raising the possibility that patterns with diversity would shift with broader geographic sampling or data from summers in other years.

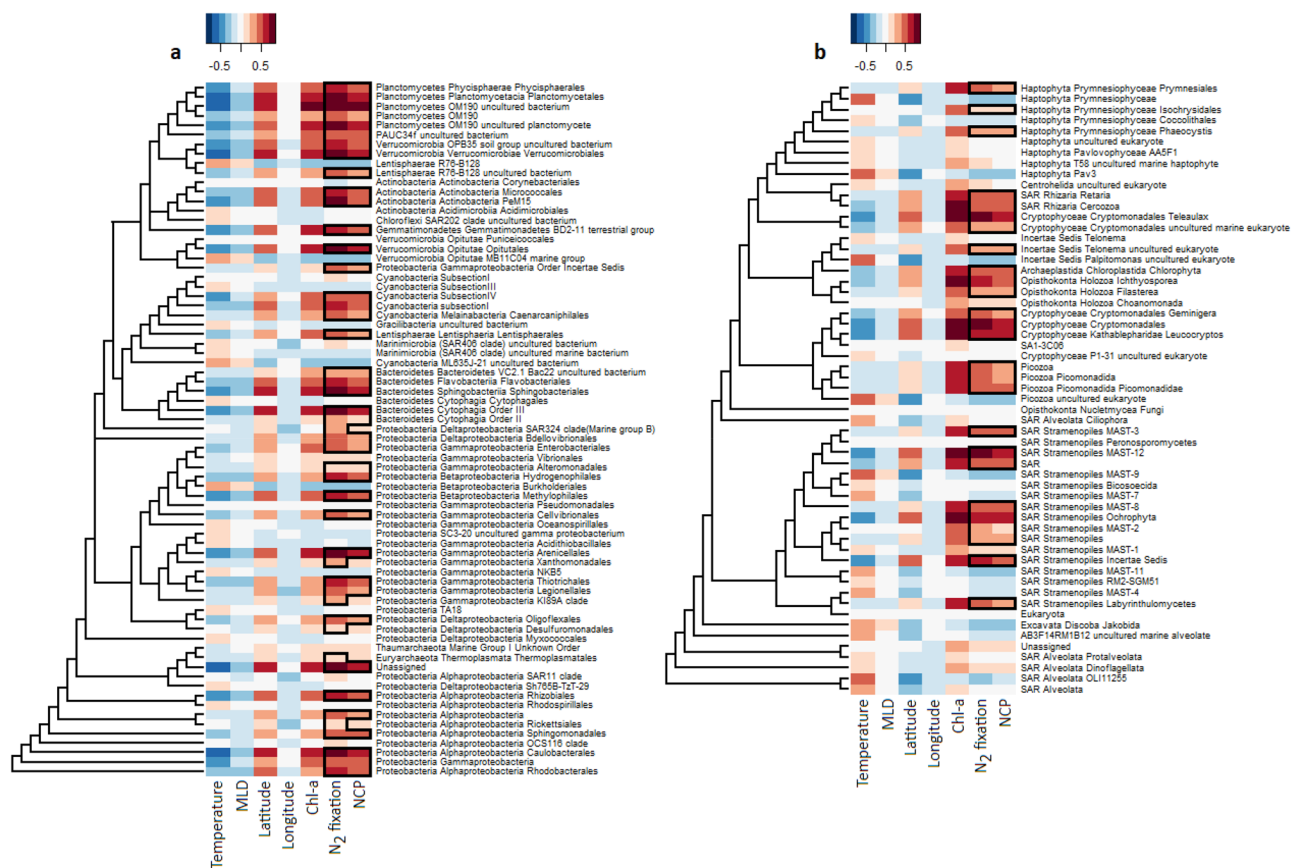


**Figure 4.** Principal Coordinates Analysis (PCoA) of (a) prokaryotic 16S and (b) eukaryotic 18S samples, ordinated using weighted Bray–Curtis dissimilarity. Samples are shaded based on whether they were collected north or south of the northern wall of the Gulf Stream in that year as determined from monthly EN4 temperature data<sup>42</sup>. Samples denoted with asterisks originate from the Mid-Atlantic Bight. Figure created in R 3.5.2 (<https://www.r-project.org/>).



**Figure 5.** Linear regressions of (a) NCP and (b) N<sub>2</sub> fixation rates vs 18S Shannon's H diversity. Linear regressions of (c) NCP and (d) N<sub>2</sub> fixation rates vs 16S Shannon's H diversity. Significance and goodness-of-fit are displayed for each plot. All diversity metrics were calculated at the OTU level. Figure created in R 3.5.2 (<https://www.r-project.org/>).





**Figure 6.** Heatmaps showing correlations between abundances of specific (a) prokaryotic 16S and (b) eukaryotic 18S taxa and measured metadata parameters for all three years of sampling, as determined using Partial Least Squares (PLS) regression analysis. The strength of correlations is illustrated by the color scale, with positive and negative correlations shown using warm and cool tones. Cells outlined in black rectangles indicate taxa with correlations to  $N_2$  fixation or NCP greater than 0.2. Taxa are binned at the fourth taxonomic rank for both 16S and 18S data. Only taxa for which the sum of corresponding reads is more than 0.5% of the total sum of all 18S or 16S reads are displayed. Dendrograms drawn to the left of each heatmap illustrate phylogenetic relationships between row taxa. Note that horizontal branch lengths have been compressed for layout reasons and no longer represent exact distances. Figure created in R 3.5.2 (<https://www.r-project.org/>).

Furthermore, our dataset remains too limited to investigate the possibility that underlying relationships are curvilinear as opposed to linear. Different taxonomic definitions, such as amplicon sequence variants<sup>60</sup> could also affect biodiversity metrics, a topic that is beyond the scope of this study but worthy of further investigation.

**Fine-scale relationships: key community members associated with  $N_2$  fixation and NCP.** Our partial least squares (PLS) regression analyses identified specific prokaryotic and eukaryotic taxa as jointly associated with high  $N_2$  fixation and high NCP rates (Fig. 6). Among prokaryotes, abundances for all *Planctomycetes* taxa binned at the fourth taxonomic rank were related to  $N_2$  fixation and NCP. 2 of 4 *Actinobacteria* taxa, 3 of 5 *Verrucomicrobia*, and 2 of 3 *Betaproteobacteria* also exhibited relationships with  $N_2$  fixation and NCP, as did *Arenicellales* (*Gammaproteobacteria*), *Sphingobacteriales*, and *Cytophagia* Order III from within *Bacteroidetes*. PLS regression analyses performed at the genus level identified a strong relationship between UCYN-A, a haptohyte-associated,  $N_2$  fixing group of unicellular cyanobacteria, and both  $N_2$  fixation and NCP (Dataset S1g).

*Planctomycetes* are largely heterotrophic, can exhibit surface and particle-associated lifestyles<sup>61</sup> and have been observed at high abundances in coastal environments<sup>62</sup>. Consequently, relationships with productivity and diazotrophic activity could potentially reflect the particle-rich environment suggested by high quantitative abundances of *Chrysophyceae* and *Aureococcus anophagefferens*. A higher concentration of phytoplankton surfaces for growth might similarly explain  $N_2$  fixation and NCP associations displayed by *Bacteroidetes* and *Verrucomicrobia*, whose marine representatives are also typically particle-associated<sup>63–65</sup>. Alternatively, recent research indicates that heterotrophic bacterial diazotrophs, including some *Planctomycetes*, are more widespread and abundant in the surface ocean than previously thought<sup>17</sup>, potentially contributing directly to  $N_2$  fixation and thereby promoting new production.

Eukaryotic groups associated with both  $N_2$  fixation and NCP included most cryptophytes, *Ichthyosporea*, *Cercozoa*, *Prymnesiales*, *Retaria*, *Chlorophyta*, and several marine stramenopiles (MASTs) (Fig. 6). The taxon



*Ochrophyta*, containing the dominant stramenopiles within the coastal transect—*Chrysophyceae* and *Aureococcus anophagefferens*—was also predictably related to NCP and  $N_2$  fixation. At the genus level, UCYN-A's hypothesized prymnesiophyte host, *Braarudosphaera bigelowii*, was related to  $N_2$  fixation and NCP (Dataset S1h). PLS regression analyses of taxonomic abundances in relation to macronutrients, trace metals, and silica are shown in the Supplementary Material.

Network and graph-alignment analyses revealed several eukaryotic taxa associated with  $N_2$  fixation and NCP as hub organisms within their subnetworks or communities. These hub taxa showed high numbers of significant co-occurrence relationships, implying that they could play central community roles (Fig. 7). For example, Subnetwork 1 in Fig. 7 described a community associated with high  $N_2$  fixation and productivity sites in the 2017 coastal bloom (Fig. S3). Within this community, network analysis identified marine stramenopiles as important hub taxa, with four stramenopile genera including two chrysophytes placing among the top 10 keystone nodes. Notably, our results highlighted *Braarudosphaera bigelowii* as the taxon with the single highest network centrality within its  $N_2$  fixation and NCP-associated subnetwork, suggesting that this diazotroph-hosting prymnesiophyte may be ecologically influential. Taxa with high network centrality are critical to subnetwork stability by exhibiting a high number of connections with other OTUs. Other marine microbiome studies along the California coast<sup>66</sup> and off Brazil<sup>67</sup> have also found evidence that *B. bigelowii* may play important ecological roles within other microbial ecosystems. Equally interesting were taxa associated with  $N_2$  fixation and NCP in PLS regression analysis but exhibiting few sub-network connections, such as cryptophyte algae and *Aureococcus*. Their elevated abundance in productive samples but low participation in network structure could potentially suggest that cryptophytes and *Aureococcus* benefit opportunistically from ecosystem conditions but possess fewer direct interactions with other microbes in their environment.

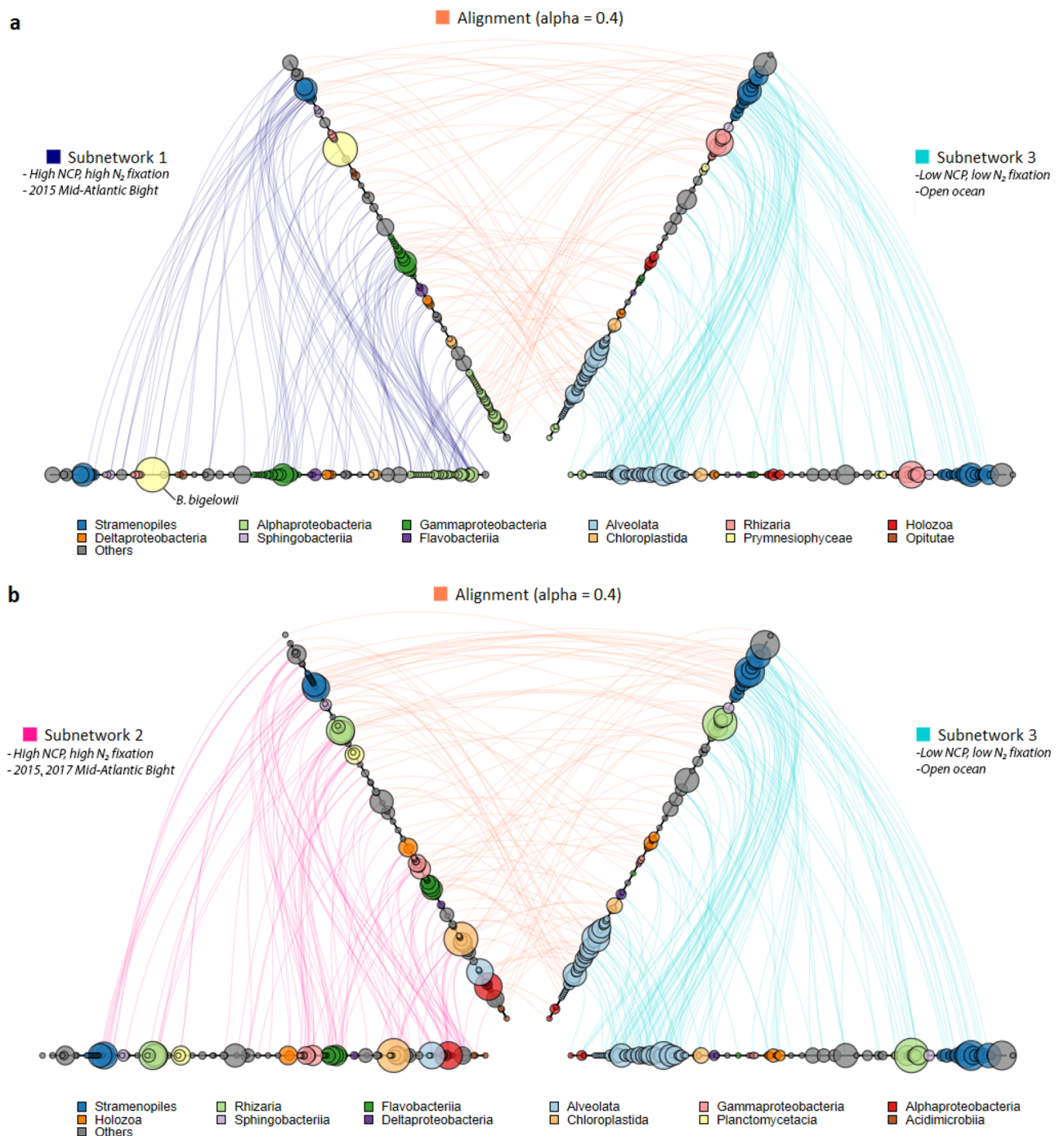
Subnetwork 2 described a community generally associated with coastal sampling in 2015 and 2017 (Fig. S3). A similar centrality analysis revealed several chlorophyte genera—*Micromonas*, *Bathycoccus*, and *Ostreococcus*—as important keystone nodes. High network centrality associated with these taxa could conceivably reflect a potential role of these small, genetically-streamlined chlorophytes as prey within the community for larger mixotrophic and heterotrophic protists and a shared response alongside other taxa to favorable environmental conditions. That said, possibilities exist for more complex interactions—members of *Micromonas* in polar waters have recently been identified as mixotrophs, practicing bacterivory<sup>68</sup>. Our data cannot directly provide insight into these hypotheses. However, the high network centrality of chlorophyte taxa and their abundance in high  $N_2$  fixation and NCP samples suggest that they may be ecologically important members of the productive coastal community sampled.

Stramenopiles, including several *Chrysophyceae*, were also identified as essential taxa within the low  $N_2$  fixation, low-productivity Subnetwork 3 together with alveolates, driving much of the network structure within this module. To emphasize common patterns between stramenopiles in these subnetworks, we now examine the results of our graph alignment analysis. A graph alignment consists of identifying OTUs from two distinct communities that possess (i) similar marker gene sequences (i.e., sequence identity) and (ii) similar roles within the communities' topological structure (i.e., patterns of centrality). Aligned OTUs from different subnetworks imply similar ecological roles in their respective communities. A large fraction of the significant alignments between Subnetwork 3 and Subnetworks 1 and 2 involved stramenopiles and alveolates from Subnetwork 3. This alignment result suggests that a community transition between these conditions could be associated with the resilience of these groups of eukaryotes and their related ecological functions.

Our overall findings corroborated this result. The observed close correspondence between peaks in  $N_2$  fixation and *Chrysophyceae* abundances supports the theory that stramenopiles potentially play a key role in our study region. However, this also highlights our lack of general understanding of the function and role of stramenopiles. To our knowledge, the oceanographic literature so far does not record any known direct association between *Chrysophyceae* and diazotrophs that could drive such a pattern, motivating the proposal of hypotheses to drive further studies. *Chrysophyceae* and diazotrophic activity could simply co-vary in response to a separate stimulatory factor. Alternatively, diazotrophic inputs of N might, through recycling via the microbial loop, act to promote *Chrysophyceae* growth. Mixotrophic *Chrysophyceae* possess diverse metabolic capabilities including heterotrophic uptake of organic matter using phagotrophy, potentially supporting such a hypothesis<sup>56,57</sup>. High abundances of *Chrysophyceae* might, on the other hand, create favorable conditions for UCYN-A and its host *B. bigelowii* through the competitive exclusion of competitors. Or, a direct interaction between *Chrysophyceae* and yet-uncharacterized  $N_2$ -fixers could represent an unlikely but intriguing possibility. At any rate, this pattern of association between *Chrysophyceae* and high measured  $N_2$ -fixing activity warrants a more detailed investigation driven by interaction studies.

## Conclusion

Our study of prokaryotic and eukaryotic community structure within hotspots of  $N_2$  fixation and NCP revealed an intriguing microbial ecology. We observed a strong negative relationship between eukaryotic diversity and both  $N_2$  fixation and  $O_2$ /Ar-derived NCP, a finding suggesting that low eukaryotic microbial diversity at productive coastal sites might potentially be driven by competitive exclusion. Notably, we observed high abundances of *Chrysophyceae* occurring in strikingly close conjunction with high  $N_2$  fixation rates and NCP, along with elevated abundances of *Aureococcus anophagefferens*, chlorophytes, cryptophytes, and *Planctomycetes*. Such a community structure indicates that observed high  $N_2$  fixation rates may be occurring in microbial ecosystems with a high potential to process and take up diverse forms of organic material and nutrients. Aided by our quantitative study design, our network analysis highlighted *Braarudosphaera bigelowii*, chlorophytes, and *Chrysophyceae* and other marine stramenopiles as potentially important community hub taxa. Overall, our findings point to notable taxa driving the ecological network structure of coastal  $N_2$  fixation hotspots and, particularly in light of the



**Figure 7.** Alignment plots from co-occurrence and graph alignment analyses illustrate differences in ecological patterns and relationships between low  $N_2$  fixation + Low NCP and high  $N_2$  fixation + high NCP subnetworks. **(a)** Visualization of network relationships within and contrasted between Subnetwork 1 (2015 MAB, high  $N_2$  fixation) and Subnetwork 3 (open ocean, low  $N_2$  fixation) **(b)** and Subnetwork 2 (2015 + 2017 MAB, high  $N_2$  fixation) and Subnetwork 3 (open ocean, low  $N_2$  fixation). The leftmost pair of identical axes represent taxa belonging to the first subnetwork, with edges between the left two axes signifying significant co-occurrence relationships between taxa within the subnetwork. The rightmost axes and lines similarly represent taxa from the second subnetwork and within-network co-occurrence relationships. Edges between the two networks represent aligned taxa, where an OTU shares a similar taxonomic identity (sequence similarity) and a high degree of network topological similarity with an OTU in the other subnetwork. Alignments suggest that OTUs fill similar ecological roles within their respective subnetworks. Circular nodes represent individual OTUs, colored and grouped at the third taxonomic rank. Larger circle diameter indicates higher network importance (centrality). Taxa are ordered by centered log-ratio (CLR) normalized abundance, with higher-abundance genera placed further from the origin. Figure produced using HiveAlign (<https://univ-nantes.io/erwan.delage/HiveAlign>).

establishment of a new LTER site along the northeast U.S. coastal shelf (NES-LTER)<sup>69</sup>, can serve as a springboard for targeted investigations into the processes mediated by these microorganisms of interest.

## Materials and methods

**Underway measurements of N<sub>2</sub> fixation rates and net community production.** Our combined set of new and previously-published data includes around 10 250 km of continuous and discrete measurements across the western North Atlantic Ocean over three summer voyages aboard the *R/V Atlantic Explorer*. Previously-reported surface layer N<sub>2</sub> fixation rates<sup>22,37</sup> were measured at high resolution using the Flow-through incubation Acetylene Reduction Assay by Cavity ring-down Spectroscopy (FARACAS)<sup>38</sup>, which utilizes a cavity ring-down spectrometer to measure the conversion of dissolved acetylene to ethylene by diazotrophs within surface seawater. Previously-reported measurements of NCP<sup>22,36,37</sup> were conducted using the dissolved O<sub>2</sub>/Ar method via Equilibrator Inlet Mass Spectrometry<sup>70</sup>. Detailed protocols and data analysis procedures can be found in previous publications<sup>22,36,37</sup>. We conducted a sensitivity analysis to assess the effect of vertical O<sub>2</sub>/Ar fluxes upon calculated NCP rates, detailed in the Supplementary Methods, which we deemed to be minimal. The cruise tracks, conducted from 3 to 13 August 2015, 3–12 August 2016, and 29 July–7 August 2017, are illustrated in Fig. 1, with associated physical and biogeochemical properties shown in Fig. S1 and S2. Our N<sub>2</sub> fixation technique calculates daily fixation rates, while NCP measurements integrate over 2–8 days, both reasonable timescales for comparisons between community structure and these rate measurements.

**Sample collection and library preparation for 16S and 18S rDNA amplicon sequencing.** Microbial community composition data from 2015 was previously published in Wang et al., 2018<sup>36</sup>, whereas molecular data and samples from the 2016 and 2017 are original to this paper. Molecular samples for rDNA amplicon sequencing were collected every 6–8 h using 0.22 µm polycarbonate filters (Millipore, Billerica, MA, USA) from surface (3–5 m) seawater collected from towfish<sup>71</sup> and 5 m CTD samples. Between 1–2 l (mean: 1.2 l) were filtered using a peristaltic pump, with smaller volumes filtered at particularly high-biomass stations. In 2015, due to the leveraging of samples initially collected for RNA sequencing, all utilized filters were preserved in RNAlater (Thermo Fisher, Waltham, MA, USA), while RNAlater preservation was not employed for 2016 and 2017 samples. Otherwise, the sampling methodology was identical across all three years. Filters were flash-frozen in liquid nitrogen following filtration and stored at –80 °C.

The addition of internal standard control sequences was performed following the protocol described in Wang et al., 2018<sup>36</sup> for all samples. For the 2016 and 2017 samples, new standard stock solutions were prepared and amounts added were adjusted based on the 2015 sequencing results. *T. thermophilus* gDNA standard was added to each sample before bead-beating in 50 µl volumes with amounts of 2,800,000 copies (3.05 ng) per sample for 2015 samples and 14,040,000 copies (15.25 ng) per sample for 2016 samples. For both 2015 and 2016 samples, 5,780,000 copies (0.68 ng) of *S. pombe* gDNA were introduced. For 2017 samples, 0.66 ng of *S. pombe* gDNA and 6.1 ng of *T. thermophilus* gDNA were added, corresponding to approximately 5,610,000 copies sample<sup>–1</sup> and 5,620,000 copies sample<sup>–1</sup> respectively.

DNA extractions, PCR amplification of the 16S rDNA V4 and the 18S rDNA V4 regions, sample quantitation, and library pooling followed the same procedure as Wang et al., 2018<sup>36</sup>. Briefly, DNA extraction was performed using the Qiagen DNeasy Plant Mini Kit, utilizing the manufacturer's instructions with slight modifications<sup>72</sup>. PCR parameters were as follows for 16S samples: an initial denaturing step of 94 °C for 3 min, followed by 30 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min, and ending with a final elongation step at 72 °C for 10 min. 18S parameters were identical apart from an annealing temperature of 57 °C. The PCR step employed the custom 16S rDNA V4 primers 515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3') and 805 R (5'-GACTACNVGGG TATCTAAT-3') and 18S rDNA V4 primers F (5'-CCAGCASCYGCAGGTAATTC-3') and R (5'-ACTTTCGTT CTTGAT-3'), with attached Illumina adapters and barcodes<sup>36</sup>. 16S PCR reactions (25 µl volume) contained 2.5 µl 10× PCR buffer, 0.5 µl dNTP mix (10 mM each), 1 µl 50 mM MgSO<sub>4</sub>, 0.5 µl of each primer (10 µM), 0.1 µl Platinum Taq Hi-Fidelity Polymerase, and 19.4 µl of sterile water, plus 0.5 µl of template. 18S PCR reaction mixtures were identical, although template and polymerase amounts were adjusted (0.1–0.2 µl polymerase, 0.5–3 µl template) with corresponding adjustments to sterile water volume to address weak amplification of some samples. PCR products were purified with the Qiagen QIAquick PCR Purification Kit, quantified using a Qubit 3.0 fluorometer, then pooled in equimolar amounts into a 16S and an 18S library. Illumina MiSeq (300 bp PE, V3) sequencing was subsequently performed at the Duke Center for Genomic and Computational Biology.

**rDNA sequencing data processing and OTU picking.** We obtained 43,258,158 paired-end reads from 167 surface samples. Paired-end sequences were trimmed, merged, quality filtered, demultiplexed, then filtered for chimeras as in Wang et al., 2018<sup>36</sup>, yielding 11,474,817 16S sequences and 4,988,755 18S sequences. Primer and other non-biological sequences were then filtered using Tagcleaner<sup>73</sup>. Open-reference OTU picking was subsequently performed in QIIME (97% similarity, Usearch 6.1<sup>74</sup>, SILVA 123.1<sup>75</sup>)<sup>36</sup> (see Supplementary Material for additional details). Following taxonomy assignment, we removed low read count samples (<100 sequences) and filtered internal standard, plastid, mitochondrial, and metazoan rDNA sequences from our taxonomy tables. Phylogenetic trees were generated using the make\_phylogeny.py script in QIIME, which utilizes the FastTree method.

**Statistical analyses.** The combined 2015–2017 sequencing dataset included 8 841 16S OTUs across 79 16S samples and 14 344 18S OTUs across 79 18S samples, post-filtering. Sequencing data were imported into R 3.4.1<sup>76</sup> using the phyloseq<sup>77</sup> package. Calculation of quantitative abundances was performed as in Satinsky et al., 2013<sup>78</sup>, with the total rDNA gene abundance l<sup>–1</sup> of a taxon determined as the number of reads mapping



to that OTU in the sequencing output, divided by the product of the standard recovery ratio (# standard reads recovered / # standard reads introduced) and the volume filtered. The internal standard recovery ratios obtained in this study are shown in (Fig. S4). Details of filtration of low-count samples and outliers and rarefaction and ordination procedures are described in the Supplementary Material.

The internal standard technique has previously demonstrated good agreement with other quantification techniques including flow cytometry counts, HPLC, phospholipid fatty acid (PLFA) analysis, and substrate-induced respiration (SIR) analysis<sup>79–81</sup>. However, the approach remains subject to limitations. This quantitative method does not account for variable rDNA gene copy number across bacterial and eukaryotic taxa. Prokaryotic rDNA copy numbers mostly vary between 1 and 15 copies per cell<sup>82</sup>, while eukaryotic rDNA gene copies cell<sup>-1</sup> can vary by orders of magnitude. Additionally, the internal standard technique assumes that recovery rates are identical for internal standard reads and natural sequences, an assumption that primer, amplification, or DNA extraction biases could impact. Consequently, calculated abundances should not be interpreted as absolute estimates of organismal cell concentrations, but rather reflect the changing quantitative abundance of taxa across samples.

Calculation of quantitative 16S rDNA gene abundances produced a discrepancy in which samples collected in 2015 displayed significantly higher total 16S rDNA abundances (Wilcoxon rank-sum test,  $p < 0.001$ ) than 16S rDNA samples obtained in 2016 and 2017. This difference in calculated abundances likely results from quantitation or dilution errors introduced while synthesizing the varying 16S spike quantities and stock solutions between years as mentioned above and from differences in preservation protocols. To investigate this, we compared our calculated SAR11 rDNA copy numbers for stations within 400 km of Bermuda against reported summer SAR11 abundances of  $1.5\text{--}2.0 \times 10^8$  cells/l for this region<sup>83,84</sup>. While calculated 2016 and 2017 SAR11 abundances ( $1.41 \times 10^8$  cells/l and  $1.38 \times 10^8$  cells/l) aligned closely with values from the literature, samples from 2015 near Bermuda displayed a median SAR11 abundance of  $6.5 \times 10^8$  cells/l. Consequently, we adjusted calculated 2015 16S rDNA abundances by a factor of 0.21. Adjusted 2015 organismal abundances for SAR11 and *Prochlorococcus* align closely with expected abundances from the literature (Table S2). We further cross-validated our corrected organismal abundances for *Prochlorococcus* against patterns of *Prochlorococcus* abundance obtained from concurrently-collected flow cytometry samples, observing good agreement in abundance patterns between the two datasets (Table S1, Fig. S5). The corrected 2015 16S rDNA dataset was subsequently used alongside the 2016 and 2017 datasets in generating taxonomy plots, ordination analyses, Partial least squares (PLS) regression analyses, and co-occurrence network construction.

Partial least squares (PLS) regression analysis was performed following Wang et al., 2018<sup>36</sup> upon 18S and 16S taxa at the fourth and sixth taxonomic ranks. The fourth taxonomic rank was the highest resolution for which heatmap figures are readily visually interpretable, while the sixth (genus) rank is the highest taxonomic level that the 18S and 16S rDNA amplicon can accurately and reliably resolve. Taxa at the fourth and sixth taxonomic ranks with read counts comprising  $< 0.5\%$  of the total sum of 18S or 16S reads were filtered from analyses. Two separate PLS regression analyses were performed: one utilizing the full molecular dataset from all 3 years to assess taxon-specific relationships with NCP and  $N_2$  fixation, and another utilizing only the community data from the subset of samples collected in 2016 and 2017 to investigate relationships between specific taxa and macronutrient and trace metal concentrations (see Supplementary Material). To handle missing data, the non-iterative partial least squares (nipals) algorithm was employed.

**Network analysis.** Weighted gene correlation network analysis (WGCNA)<sup>85</sup> was performed to select sub-networks of co-occurring marine microbial OTUs from the whole dataset. Following the WGCNA standard protocol<sup>86</sup>, we constructed an adjacency matrix of the microbial community dataset using Pearson correlations ( $R^2$  cutoff = 0.85)<sup>87</sup>. A soft-thresholding power law ( $p = 8$ ) was applied to fit the weighted graph into a scale-free topology. A topological overlap measure (TOM) was then calculated for each pair of OTUs, based on the weight of their pairwise correlations with close neighbors and weighted correlations with other OTUs. Finally, hierarchical clustering based on TOM was performed to identify sub-networks of co-occurring taxa. Two sub-networks (Subnetworks 1 and 2) that were most significantly positively associated with both NCP and  $N_2$  fixation (Figs. S3, S6) were extracted, as well as the subnetwork (Subnetwork 3) most negatively correlated with these rates, although not significantly.

A global, robust network was built via FlashWeave<sup>88</sup>. Induced subgraphs, as identified from WGCNA, were then extracted. Subnetwork comparison was then performed using L\_GRAAL, a graph alignment tool<sup>89</sup>. L\_GRAAL aligns OTUs from two distinct co-occurrence networks when they share (i) similar topological properties and (ii) rDNA sequence homology ( $\alpha = 0.4$  for accurate consensus between both features)<sup>90</sup>. Visualization of subnetworks and alignments was performed with HiveAlign. More details on specific steps taken for WGCNA, Flashweave, and L-GRAAL analyses can be found in the Supplementary Material.

Received: 26 August 2020; Accepted: 27 January 2021

## References

1. Falkowski, P. G. Evolution of the nitrogen cycle and its influence on the biological sequestration of CO<sub>2</sub> in the ocean. *Nature* **387**, 272–275. <https://doi.org/10.1038/387272a0> (1997).
2. Boyd, P. W. & Doney, S. C. in *Ocean Biogeochemistry: The Role of the Ocean Carbon Cycle in Global Change* (ed. Michael J. R. Fasham) 157–193 (Springer, 2003).
3. Keith Moore, J., Doney, S. C., Lindsay, K., Mahowald, N. & Michaels, A. F. Nitrogen fixation amplifies the ocean biogeochemical response to decadal timescale variations in mineral dust deposition. *Tellus B Chem. Phys. Meteorol.* **58**, 560–572. <https://doi.org/10.1111/j.1600-0889.2006.00209.x> (2006).

4. Mulholland, M. R. *et al.* High rates of N<sub>2</sub> fixation in temperate, western North Atlantic coastal waters expand the realm of marine diazotrophy. *Global Biogeochem. Cycles* **33**, 826–840. <https://doi.org/10.1029/2018gb006130> (2019).
5. Zehr, J. P. & Capone, D. G. Changing perspectives in marine nitrogen fixation. *Science* **368**, eaay9514, <https://doi.org/10.1126/science.aay9514> (2020).
6. Zehr, J. P. & Kudela, R. M. Nitrogen cycle of the open ocean: From genes to ecosystems. *Ann. Rev. Mar. Sci.* **3**, 197–225. <https://doi.org/10.1146/annurev-marine-120709-142819> (2011).
7. Langlois, R. J., Mills, M. M., Ridame, C., Croot, P. & LaRoche, J. Diazotrophic bacteria respond to Saharan dust additions. *Mar. Ecol. Prog. Ser.* **470**, 1–14. <https://doi.org/10.3354/meps10109> (2012).
8. Mark Moore, C. *et al.* Large-scale distribution of Atlantic nitrogen fixation controlled by iron availability. *Nat. Geosci.* **2**, 867. <https://doi.org/10.1038/ngeo667> (2009).
9. Mills, M. M., Ridame, C., Davey, M., La Roche, J. & Geider, R. J. Iron and phosphorus co-limit nitrogen fixation in the eastern tropical North Atlantic. *Nature* **429**, 292–294. <https://doi.org/10.1038/nature02550> (2004).
10. Turk-Kubo, K. A., Karamchandani, M., Capone, D. G. & Zehr, J. P. The paradox of marine heterotrophic nitrogen fixation: Abundances of heterotrophic diazotrophs do not account for nitrogen fixation rates in the Eastern Tropical South Pacific. *Environ. Microbiol.* **16**, 3095–3114. <https://doi.org/10.1111/1462-2920.12346> (2014).
11. Dugenne, M., Henderikx Freitas, F., Wilson, S. T., Karl, D. M. & White, A. E. Life and death of *Crocospira* sp. in the Pacific Ocean: Fine scale predator–prey dynamics. *Limnol. Oceanogr.* <https://doi.org/10.1002/lno.11473> (2020).
12. O’Neil, J. M., Metzler, P. M. & Glibert, P. M. Ingestion of 15N<sub>2</sub>-labelled *Trichodesmium* spp. and ammonium regeneration by the harpacticoid copepod *Macrosetella gracilis*. *Mar. Biol.* **125**, 89–96. <https://doi.org/10.1007/BF00350763> (1996).
13. Wang, W.-L., Moore, J. K., Martiny, A. C. & Primeau, F. W. Convergent estimates of marine nitrogen fixation. *Nature* **566**, 205–211. <https://doi.org/10.1038/s41586-019-0911-2> (2019).
14. Hagino, K., Onuma, R., Kawachi, M. & Horiguchi, T. Discovery of an endosymbiotic nitrogen-fixing cyanobacterium UCYN-A in *Braarudosphaera bigelowii* (Prymnesiophyceae). *PLoS ONE* **8**, e81749. <https://doi.org/10.1371/journal.pone.0081749> (2013).
15. Thompson, A. W. *et al.* Unicellular cyanobacterium symbiotic with a single-celled eukaryotic alga. *Science* **337**, 1546 (2012).
16. Bombar, D., Paerl, R. W. & Riemann, L. Marine non-cyanobacterial diazotrophs: Moving beyond molecular detection. *Trends Microbiol.* **24**, 916–927. <https://doi.org/10.1016/j.tim.2016.07.002> (2016).
17. Delmont, T. O. *et al.* Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nat. Microbiol.* **3**, 804. <https://doi.org/10.1038/s41564-018-0176-9> (2018).
18. Moisan, P. H., Serros, T., Paerl, R. W., Beinart, R. A. & Zehr, J. P. Gammaproteobacterial diazotrophs and nifH gene expression in surface waters of the South Pacific Ocean. *ISME J.* **8**, 1962–1973. <https://doi.org/10.1038/ismej.2014.49> (2014).
19. Knapp, A. The sensitivity of marine N<sub>2</sub> fixation to dissolved inorganic nitrogen. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2012.00374> (2012).
20. Shiozaki, T. *et al.* Diazotroph community structure and the role of nitrogen fixation in the nitrogen cycle in the Chukchi Sea (western Arctic Ocean). *Limnol. Oceanogr.* <https://doi.org/10.1002/lno.10933> (2018).
21. Inomura, K., Bragg, J., Riemann, L. & Follows, M. J. A quantitative model of nitrogen fixation in the presence of ammonium. *PLoS ONE* **13**, e0208282. <https://doi.org/10.1371/journal.pone.0208282> (2018).
22. Tang, W. *et al.* Revisiting the distribution of oceanic N<sub>2</sub> fixation and estimating diazotrophic contribution to marine production. *Nat. Commun.* **10**, 831. <https://doi.org/10.1038/s41467-019-08640-0> (2019).
23. Mulholland, M. R. *et al.* Rates of dinitrogen fixation and the abundance of diazotrophs in North American coastal waters between Cape Hatteras and Georges Bank. *Limnol. Oceanogr.* **57**, 1067–1083. <https://doi.org/10.4319/lo.2012.57.4.1067> (2012).
24. Harding, K. *et al.* Symbiotic unicellular cyanobacteria fix nitrogen in the Arctic Ocean. *Proc. Natl. Acad. Sci.* **115**, 13371. <https://doi.org/10.1073/pnas.1813658115> (2018).
25. Fonseca-Batista, D. *et al.* Evidence of high N<sub>2</sub> fixation rates in the temperate northeast Atlantic. *Biogeosciences* **16**, 999–1017. <https://doi.org/10.5194/bg-16-999-2019> (2019).
26. Frischkorn, K. R., Rouco, M., Van Mooy, B. A. S. & Dyhrman, S. T. Epibionts dominate metabolic functional potential of *Trichodesmium* colonies from the oligotrophic ocean. *ISME J.* **11**, 2090–2101. <https://doi.org/10.1038/ismej.2017.74> (2017).
27. Gradoville, M. R., Crump, B. C., Letelier, R. M., Church, M. J. & White, A. E. Microbiome of *Trichodesmium* colonies from the North Pacific subtropical gyre. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2017.01122> (2017).
28. Basu, S., Gledhill, M., de Beer, D., Prabhu Matondkar, S. G. & Shaked, Y. Colonies of marine cyanobacteria *Trichodesmium* interact with associated bacteria to acquire iron from dust. *Commun. Biol.* **2**, 284. <https://doi.org/10.1038/s42003-019-0534-z> (2019).
29. Farnelid, H. *et al.* Diverse diazotrophs are present on sinking particles in the North Pacific Subtropical Gyre. *ISME J.* **13**, 170–182. <https://doi.org/10.1038/s41396-018-0259-x> (2019).
30. Schoffelen, N. J. *et al.* Phosphate availability affects fixed nitrogen transfer from diazotrophs to their epibionts. *ISME J.* **13**, 2701–2713. <https://doi.org/10.1038/s41396-019-0453-5> (2019).
31. Deng, L., Cheung, S. & Liu, H. Protist Grazers increase grazing on unicellular cyanobacteria diazotroph at night. *Front. Mar. Sci.* <https://doi.org/10.3389/fmars.2020.00135> (2020).
32. Cairns, J., Coloma, S., Sivonen, K. & Hiltunen, T. Evolving interactions between diazotrophic cyanobacterium and phage mediate nitrogen release and host competitive ability. *R. Soc. Open Sci.* **3**, 160839. <https://doi.org/10.1098/rsos.160839> (2016).
33. Jennifer, H., Just, C. Å. N. & Ivan, V. Eelgrass *Zostera marina* loss in temperate estuaries: Relationship to land-derived nitrogen loads and effect of light limitation imposed by algae. *Mar. Ecol. Prog. Ser.* **247**, 59–73 (2003).
34. Kang, Y. & Gobler, C. J. The brown tide algae, *Aureococcus anophagefferens* and *Aureoumbra lagunensis* (Pelagophyceae), allelopathically inhibit the growth of competing microalgae during harmful algal blooms. *Limnol. Oceanogr.* **63**, 985–1003. <https://doi.org/10.1002/lno.10714> (2018).
35. Frischkorn, K. R., Rouco, M., Van Mooy, B. A. S. & Dyhrman, S. T. The *Trichodesmium* microbiome can modulate host N<sub>2</sub> fixation. *Limnol. Oceanogr. Lett.* **3**, 401–408. <https://doi.org/10.1002/lo2.10092> (2018).
36. Wang, S., Lin, Y., Gifford, S., Eveleth, R. & Cassar, N. Linking patterns of net community production and marine microbial community structure in the western North Atlantic. *ISME J.* <https://doi.org/10.1038/s41396-018-0163-4> (2018).
37. Tang, W. *et al.* New insights into the distributions of nitrogen fixation and diazotrophs revealed by high-resolution sensing and sampling methods. *ISME J.* <https://doi.org/10.1038/s41396-020-0703-6> (2020).
38. Cassar, N., Tang, W. Y., Gabathuler, H. & Huang, K. Method for high frequency underway N<sub>2</sub> fixation measurements: Flow-through incubation acetylene reduction assays by cavity ring down laser absorption spectroscopy (FARACAS). *Anal. Chem.* **90**, 2839–2851. <https://doi.org/10.1021/acs.analchem.7b04977> (2018).
39. Guillou, L. *et al.* Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ. Microbiol.* **10**, 3349–3365. <https://doi.org/10.1111/j.1462-2920.2008.01731.x> (2008).
40. Not, F., Gausling, R., Azam, F., Heidelberg, J. F. & Worden, A. Z. Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ. Microbiol.* **9**, 1233–1252. <https://doi.org/10.1111/j.1462-2920.2007.01247.x> (2007).
41. Romari, K. & Vault, D. Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol. Oceanogr.* **49**, 784–798. <https://doi.org/10.4319/lo.2004.49.3.0784> (2004).
42. Good, S. A., Martin, M. J. & Rayner, N. A. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Oceans* **118**, 6704–6716. <https://doi.org/10.1002/2013jc009067> (2013).

43. Mulholland, M. R., Gobler, C. J. & Lee, C. Peptide hydrolysis, amino acid oxidation, and nitrogen uptake in communities seasonally dominated by *Aureococcus anophagefferens*. *Limnol. Oceanogr.* **47**, 1094–1108. <https://doi.org/10.4319/lo.2002.47.4.1094> (2002).
44. Hardin, G. The competitive exclusion principle. *Science* **131**, 1292 (1960).
45. Thingstad, T. F. & Lignell, R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat. Microb. Ecol.* **13**, 19–27. <https://doi.org/10.3354/ame013019> (1997).
46. Vallina, S. M. *et al.* Global relationship between phytoplankton diversity and productivity in the ocean. *Nat. Commun.* <https://doi.org/10.1038/ncomms5299> (2014).
47. Irigoien, X., Huisman, J. & Harris, R. P. Global biodiversity patterns of marine phytoplankton and zooplankton. *Nature* **429**, 863–867. <https://doi.org/10.1038/nature02593> (2004).
48. Li, W. K. W. Macroecological patterns of phytoplankton in the northwestern North Atlantic Ocean. *Nature* **419**, 154–157. <https://doi.org/10.1038/nature00994> (2002).
49. Smith, V. H. Microbial diversity-productivity relationships in aquatic ecosystems. *FEMS Microbiol. Ecol.* **62**, 181–186. <https://doi.org/10.1111/j.1574-6941.2007.00381.x> (2007).
50. Olli, K., Paerl, H. W. & Klais, R. Diversity of coastal phytoplankton assemblages Cross - ecosystem comparison. *Estuar. Coast. Shelf Sci.* **162**, 110–118. <https://doi.org/10.1016/j.ecss.2015.03.015> (2015).
51. Cermeno, P. *et al.* Species richness in marine phytoplankton communities is not correlated to ecosystem productivity. *Mar. Ecol. Prog. Ser.* **488**, 1–9 (2013).
52. Bhadury, P. & Wright, P. C. Exploitation of marine algae: biogenic compounds for potential antifouling applications. *Planta* **219**, 561–578. <https://doi.org/10.1007/s00425-004-1307-5> (2004).
53. Blom, J. F. & Pernthaler, J. Antibiotic effects of three strains of chrysophytes (*Ochromonas*, *Poterioochromonas*) on freshwater bacterial isolates. *FEMS Microbiol. Ecol.* **71**, 281–290. <https://doi.org/10.1111/j.1574-6941.2009.00800.x> (2009).
54. Gobler, C. J. *et al.* Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4352–4357. <https://doi.org/10.1073/pnas.1016106108> (2011).
55. Gobler, C. J., Lonsdale, D. J. & Boyer, G. L. A review of the causes, effects, and potential management of harmful brown tide blooms caused by *Aureococcus anophagefferens* (Hargraves et Sieburth). *Estuaries* **28**, 726–749. <https://doi.org/10.1007/bf02732911> (2005).
56. Caron, D. A. *et al.* Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nat. Rev. Microbiol.* **15**, 6–20. <https://doi.org/10.1038/nrmicro.2016.160> (2017).
57. Wilken, S., Huisman, J., Naus-Wiezer, S. & Van Donk, E. Mixotrophic organisms become more heterotrophic with rising temperature. *Ecol. Lett.* **16**, 225–233. <https://doi.org/10.1111/ele.12033> (2013).
58. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME J.* **11**, 2639. <https://doi.org/10.1038/ismej.2017.119> (2017).
59. Rosselló-Mora, R. & Amann, R. The species concept for prokaryotes. *FEMS Microbiol. Rev.* **25**, 39–67. <https://doi.org/10.1111/j.1574-6976.2001.tb00571.x> (2001).
60. Needham, D. M., Sachdeva, R. & Fuhrman, J. A. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J.* **11**, 1614–1629. <https://doi.org/10.1038/ismej.2017.29> (2017).
61. Faria, M. *et al.* Planctomycetes attached to algal surfaces: Insight into their genomes. *Genomics* **110**, 231–238. <https://doi.org/10.1016/j.ygeno.2017.10.007> (2018).
62. Pizzetti, I. *et al.* Temporal variability of coastal planctomycetes clades at Kabeltonne Station, North Sea. *Appl. Environ. Microbiol.* **77**, 5009–5017. <https://doi.org/10.1128/aem.02931-10> (2011).
63. Yilmaz, P., Yarza, P., Rapp, J. Z. & Glöckner, F. O. Expanding the world of marine bacterial and archaeal clades. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2015.01524> (2016).
64. Buchan, A., LeCleir, G. R., Gulvik, C. A. & Gonzalez, J. M. Master recyclers: Features and functions of bacteria associated with phytoplankton blooms. *Nat. Rev. Microbiol.* **12**, 686–698. <https://doi.org/10.1038/nrmicro3326> (2014).
65. Freitas, S. *et al.* Global distribution and diversity of marine Verrucomicrobia. *ISME J.* **6**, 1499. <https://doi.org/10.1038/ismej.2012.3> (2012).
66. Needham, D. M. *et al.* Dynamics and interactions of highly resolved marine plankton via automated high-frequency sampling. *ISME J.* **12**, 2417–2432. <https://doi.org/10.1038/s41396-018-0169-y> (2018).
67. Górikas Ribeiro, C., Lopes dos Santos, A., Marie, D., Pereira Brandini, F. & Vaulot, D. Small eukaryotic phytoplankton communities in tropical waters off Brazil are dominated by symbioses between Haptophyta and nitrogen-fixing cyanobacteria. *ISME J.* **12**, 1360–1374. <https://doi.org/10.1038/s41396-018-0050-z> (2018).
68. McKie-Krisberg, Z. M. & Sanders, R. W. Phagotrophy by the picoeukaryotic green alga micromonas: Implications for Arctic Oceans. *ISME J.* **8**, 1953–1961. <https://doi.org/10.1038/ismej.2014.16> (2014).
69. Northeast US Shelf LTER, <https://web.whoi.edu/nes-lter/> (2018).
70. Cassar, N. *et al.* Continuous high-frequency dissolved O<sub>2</sub>/Ar measurements by equilibrator inlet mass spectrometry. *Anal. Chem.* **81**, 1855–1864. <https://doi.org/10.1021/ac802300u> (2009).
71. Bruland, K. W., Rue, E. L., Smith, G. J. & DiTullio, G. R. Iron, macronutrients and diatom blooms in the Peru upwelling regime: brown and blue waters of Peru. *Mar. Chem.* **93**, 81–103. <https://doi.org/10.1016/j.marchem.2004.06.011> (2005).
72. Moisan, P. H., Beinart, R. A., Voss, M. & Zehr, J. P. Diversity and abundance of diazotrophic microorganisms in the South China Sea during intermonsoon. *ISME J.* **2**, 954–967. <https://doi.org/10.1038/ismej.2008.51> (2008).
73. Schmieder, R., Lim, Y. W., Rohwer, F. & Edwards, R. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *Bmc Bioinform.* <https://doi.org/10.1186/1471-2105-11-341> (2010).
74. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461> (2010).
75. Pruesse, E. *et al.* SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35**, 7188–7196. <https://doi.org/10.1093/nar/gkm864> (2007).
76. R: A language and environment for statistical computing. (R Foundation for Statistical Computing, 2017).
77. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0061217> (2013).
78. Satinsky, B. M., Gifford, S. M., Crump, B. C. & Moran, M. A. Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Microb. Metagenom. Metatranscriptom. Metaproteom.* **531**, 237–250. <https://doi.org/10.1016/b978-0-12-407863-5.00012-5> (2013).
79. Smets, W. *et al.* A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. *Soil Biol. Biochem.* **96**, 145–151. <https://doi.org/10.1016/j.soilbio.2016.02.003> (2016).
80. Lin, Y., Gifford, S., Ducklow, H., Schofield, O. & Cassar, N. Towards quantitative microbiome community profiling using internal standards. *Appl. Environ. Microbiol.* **85**, e02634-e12618. <https://doi.org/10.1128/AEM.02634-18> (2019).
81. Tkacz, A., Hortal, M. & Poole, P. S. Absolute quantitation of microbiota abundance in environmental samples. *Microbiome* **6**, 110. <https://doi.org/10.1186/s40168-018-0491-7> (2018).
82. Větrovský, T. & Baldrian, P. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS ONE* **8**, e57923. <https://doi.org/10.1371/journal.pone.0057923> (2013).
83. Carlson, C. A. *et al.* Seasonal dynamics of SAR11 populations in the euphotic and mesopelagic zones of the northwestern Sargasso Sea. *ISME J.* **3**, 283–295 (2008).



84. Morris, R. M. *et al.* SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420**, 806–810. <https://doi.org/10.1038/nature01240> (2002).
85. Langfelder, P. & Horvath, S. Eigengene networks for studying the relationships between co-expression modules. *Bmc Syst. Biol.* <https://doi.org/10.1186/1752-0509-1-54> (2007).
86. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465. <https://doi.org/10.1038/nature16942> (2016).
87. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559. <https://doi.org/10.1186/1471-2105-9-559> (2008).
88. Tackmann, J., Rodrigues, J. F. M. & von Mering, C. Rapid inference of direct interactions in large-scale ecological networks from heterogeneous microbial sequencing data. *bioRxiv* (2018).
89. Malod-Dognin, N. & Pržulj, N. L-GRAAL: Lagrangian graphlet-based network aligner. *Bioinformatics* **31**, 2182–2189. <https://doi.org/10.1093/bioinformatics/btv130> (2015).
90. Mandakovic, D. *et al.* Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience. *Sci. Rep.* <https://doi.org/10.1038/s41598-018-23931-0> (2018).

## Acknowledgements

This research was supported by an NSF-CAREER grant awarded to NC (#1350710) a Chateaubriand Fellowship awarded to SW, and an ANR grant awarded to HP (ANR-12-PDOC-0025-01). NC, HW, and AGG were also supported by the "Laboratoire d'Excellence" LabexMER (ANR-10-LABX-19) and co-funded by the French government under the program "Investissements d'Avenir". ED was supported by ProBioSTIC-AtlanSTIC2020. AGG also thanks the postdoctoral grant from the Universidad de Las Palmas de Gran Canaria. WT was supported by several Duke Graduate School Fellowships including The Dissertation Research Travel Award: International and the Graduate Student Training Enhancement Grant. We thank the staff of BIOS as well as the crew of the R/V Atlantic Explorer. We thank Rachel Eveleth and Yajuan Lin for assistance with sample collection and measurements during expeditions. We are also thankful to Rod Johnson, Bruce Williams, and Natasha McDonald at BIOS for cruise assistance and sample analyses, and Heidi Hirsch and the R/V Baseline Explorer's crew for help with equipment shipping. We thank Zackary Johnson for assistance with running and analyzing flow cytometry samples. We are also grateful to Nicholas Foukal, Roberta Hamme, Brivaëla Moriceau, Hans Gabathuler, Geoffrey Smith, Adeline Bidault, Jacqueline Robinson-Hamm, Adrienne Pittman, Jennifer Wernegreen, and Heather Hemric.

## Author contributions

S.W. conducted core writing of the manuscript and molecular sampling, library preparation, sequence analysis, and downstream analyses with input from S.G. W.T. performed  $N_2$  fixation measurements and data analysis. E.D., S.W., and D.E. conducted network and graph alignment analyses. H.W. and A.G. performed trace metal sampling and analysis. N.C. and H.P. conceived of the project and provided supervision. All authors participated in and provided feedback on writing, editing, and revision of the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-84969-1>.

**Correspondence** and requests for materials should be addressed to N.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021